

What is claimed is:

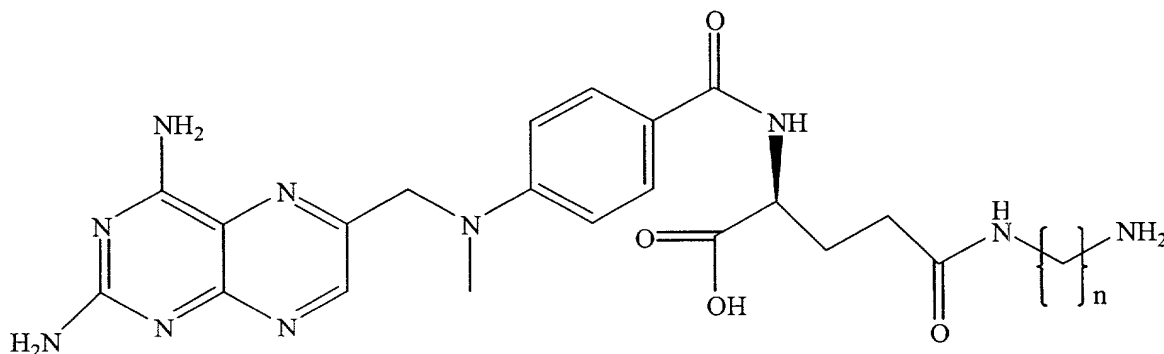
1. A method for identifying which protein from a pool of candidate proteins catalyzes in a cell a bond forming reaction between a first substrate and a second substrate, comprising:
 - (a) providing a dimeric small molecule which comprises a known moiety that binds a known receptor domain covalently linked with a moiety that contains the first substrate;
 - (b) introducing the dimeric molecule into a cell which comprises
 - i) a first fusion protein comprising the known receptor domain,
 - ii) a second fusion protein comprising the second substrate,
 - iii) a protein from the pool of candidate proteins, and
 - iv) a reporter gene wherein expression of the reporter gene is conditioned on the proximity of the first fusion protein to the second fusion protein;
 - (c) permitting the dimeric molecule to bind to the first fusion protein and to enzymatically form a bond with the second fusion protein so as to activate the expression of the reporter gene;
 - (d) selecting which cell expresses the reporter gene; and
 - (e) identifying the protein that catalyzes the bond formation reaction in the cell between the first substrate and the second substrate.
2. The method of claim 1, wherein the protein is encoded by a DNA from the group consisting of genomic DNA, cDNA and synthetic DNA.
3. The method of claim 1, wherein the pool of candidate proteins is obtained by combinatorial techniques.

4. The method of claim 1, wherein the steps (b)-(e) of the method are iteratively repeated in the presence of a preparation of random proteins for competitive enzymatic bond formation so as to identify a protein having enhanced enzymatic activity.
5. The method of claim 1, wherein the cell is an insect cell, a yeast cell, a bacterial cell, or a mammalian cell.
6. The method of claim 1, wherein the cell is a yeast cell.
7. The method of claim 1, wherein the first fusion protein further comprises a DNA binding domain, and the second fusion protein further comprises a transcription activation domain.
8. The method of claim 1, wherein the first fusion protein further comprises a transcription activation domain, and the second fusion protein further comprises a DNA binding domain.
9. The method of claim 7 or 8, wherein the DNA-binding domain is LexA, Gal4 or VP16.
10. The method of claim 7 or 8, wherein the transcription activation domain is B42.
11. The method of claim 1, wherein the known moiety that binds a known receptor domain is a Methotrexate moiety, a dexamethasone moiety, FK506 moiety, an FK506 analog, a tetracycline moiety, or a cephem moiety.
12. The method of claim 1, wherein the known receptor domain is that of dihydrofolate reductase ("DHFR"), glucocorticoid receptor, FKBP12, FKBP mutants, tetracycline repressor, or a penicillin binding protein.

13. The method of claim 12, wherein the DHFR is the *E.coli* DHFR ("eDHFR").
14. The method of claim 1, wherein the first fusion protein is eDHFR-LexA or R61-LexA.
15. The method of claim 1, wherein the first fusion protein is eDHFR-B42 or R61-B42.
16. The method of claim 1, wherein the reporter gene is *Lac Z*, *ura 3*, GFP, β -lactamase, luciferase or an antibody coding region.
17. The method of claim 1, wherein the reporter gene is *Lac Z*.
18. The method of claim 1, wherein the first substrate is an amine.
19. The method of claim 1, wherein the second substrate is an amine.
20. The method of claim 1, wherein the second substrate is an amino acid sequence containing a lysine.
21. The method of claim 1, wherein the second substrate is an amino acid sequence containing a glutamine.
22. The method of claim 1, wherein the second substrate is an amino acid sequence containing -leucine-glycine-glutamine-glycine-.
23. The method of claim 1, wherein the second substrate is an amino acid sequence containing -leucine-glutamine-glycine-glycine-.

24. The method of claim 1, wherein the second substrate is an amino acid sequence containing -leucine-leucine-glutamine-glycine-.
25. The method of claim 1, wherein the second substrate is a modified staphylococcal nuclease ("SNase") or a modified thioredoxin containing an amino acid sequence containing a glutamine.
26. The method of claim 1, wherein the protein that catalyzes bond formation is a transglutaminase.
27. The method of claim 1, wherein the protein that catalyzes bond formation is a microbial transglutaminase, a tissue transglutaminase, or Factor XIII_A.

28. The method of claim 1, wherein the dimeric small molecule has the structure:



wherein n is an integer from 1 to 20.

29. The method of claim 28, wherein n is an integer from 2 to 12.
30. The method of claim 28, wherein n is an integer from 3 to 9.
31. The method of claim 28, wherein n is 5.
32. A new protein cloned by the method of claim 1.
33. A method for identifying which substrate from a pool of candidate substrates is selected in a cell by a known enzyme for a bond forming reaction between the substrate and a known amino acid, comprising:
- (a) providing a dimeric small molecule which comprises the substrate covalently linked to a moiety known to bind a receptor domain;
 - (b) introducing the dimeric molecule into a cell which comprises
 - i) a first fusion protein comprising the receptor domain,
 - ii) a second fusion protein comprising the known amino acid,

iii) the known enzyme, and

iv) a reporter gene wherein expression of the reporter gene is conditioned on the proximity of the first fusion protein to the second fusion protein;

(c) permitting the dimeric molecule to bind to the first fusion protein and to enzymatically form a bond with the second fusion protein so as to activate the expression of the reporter gene;

(d) selecting which cell expresses the reporter gene; and

(e) identifying the substrate selected by the known enzyme in the cell for the bond forming reaction between the substrate and the known amino acid.

34. The method of claim 33, the pool of candidate substrates is obtained by combinatorial techniques.

35. The method of claim 33, wherein the steps (b)-(e) of the method are iteratively repeated in the presence of a preparation of random substrates for competitive enzymatic bond formation so as to identify a substrate competitively selected by the known enzyme.

36. The method of claim 33, wherein the cell is an insect cell, a yeast cell, a bacterial cell, or a mammalian cell.

37. The method of claim 33, wherein the cell is a yeast cell.

38. The method of claim 33, wherein the first fusion protein further comprises a DNA binding domain, and the second fusion protein further comprises a transcription activation domain.

39. The method of claim 33, wherein the first fusion protein further comprises a transcription activation domain, and the second fusion protein further comprises a DNA binding domain.

40. The method of claim 38 or 39, wherein the DNA-binding domain is LexA, Gal4 or VP16.
41. The method of claim 38 or 39, wherein the transcription activation domain is B42.
42. The method of claim 33, wherein the moiety known to bind a receptor domain is a Methotrexate moiety, a dexamethasone moiety, FK506 moiety, an FK506 analog, a tetracycline moiety, or a cephem moiety.
43. The method of claim 33, wherein the receptor domain is that of dihydrofolate reductase ("DHFR"), glucocorticoid receptor, FKBP12, FKBP mutants, tetracycline repressor, or a penicillin binding protein.
44. The method of claim 43, wherein the DHFR is the *E.coli* DHFR ("eDHFR").
45. The method of claim 33, wherein the first fusion protein is eDHFR-LexA or R61-LexA.
46. The method of claim 33, wherein the first fusion protein is eDHFR-B42 or R61-B42.
47. The method of claim 33, wherein the reporter gene is *Lac Z*, *ura 3*, GFP, β -lactamase, luciferase or an antibody coding region.
48. The method of claim 33, wherein the reporter gene is *Lac Z*.
49. The method of claim 33, wherein the enzyme that catalyzes bond formation is a transglutaminase.

50. The method of claim 33, wherein the enzyme that catalyzes bond formation is a microbial transglutaminase, a tissue transglutaminase, or Factor XIIIa.

51. A transgenic cell comprising

(a) a dimeric small molecule which comprises a moiety known to bind a receptor domain covalently linked to a first substrate of an enzyme;

(b) nucleotide sequences which upon transcription encode

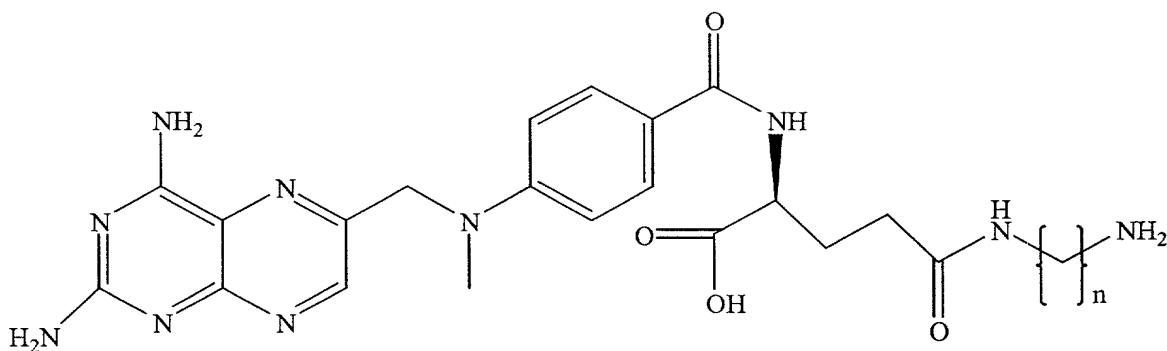
i) the enzyme,

ii) a first fusion protein comprising the receptor domain, and

ii) a second fusion protein comprising a second substrate of the enzyme; and

(c) a reporter gene wherein expression of the reporter gene is conditioned on the proximity of the first fusion protein to the second fusion protein.

52. The cell of claim 51, wherein the dimeric small molecule has the structure:



wherein n is an integer from 1 to 20.

53. The cell of claim 52, wherein n is an integer from 2 to 12.

54. The cell of claim 52, wherein n is an integer from 3 to 9.

55. The cell of claim 52, wherein n is 5.
56. The cell of claim 51, wherein the cell is an insect cell, a yeast cell, a bacterial cell, or a mammalian cell.
57. The cell of claim 51, wherein the cell is a yeast cell.
58. The cell of claim 51, wherein the first fusion protein further comprises a DNA binding domain, and the second fusion protein further comprises a transcription activation domain.
59. The cell of claim 51, wherein the first fusion protein further comprises a transcription activation domain, and the second fusion protein further comprises a DNA binding domain.
60. The cell of claim 58 or 59, wherein the DNA-binding domain is LexA, Gal4 or VP16.
61. The cell of claim 58 or 59, wherein the transcription activation domain is B42.
62. The cell of claim 51, wherein the moiety known to bind a receptor domain is a Methotrexate moiety, a dexamethasone moiety, FK506 moiety, an FK506 analog, a tetracycline moiety, or a cephem moiety.
63. The cell of claim 51, wherein the known receptor domain is that of dihydrofolate reductase ("DHFR"), glucocorticoid receptor, FKBP12, FKBP mutants, tetracycline repressor, or a penicillin binding protein.
64. The cell of claim 63, wherein the DHFR is the *E.coli* DHFR ("eDHFR").

65. The cell of claim 51, wherein the first fusion protein is eDHFR-LexA or R61-LexA.
66. The cell of claim 51, wherein the first fusion protein is eDHFR-B42 or R61-B42.
67. The cell of claim 51, wherein the reporter gene is *Lac Z*, *ura 3*, GFP, β -lactamase, luciferase or an antibody coding region.
68. The cell of claim 51, wherein the reporter gene is *Lac Z*.
69. The cell of claim 51, wherein the first substrate is an amine.
70. The cell of claim 51, wherein the second substrate is an amine.
71. The cell of claim 51, wherein the second substrate is an amino acid sequence containing a lysine.
72. The cell of claim 51, wherein the second substrate is an amino acid sequence containing a glutamine.
73. The cell of claim 51, wherein the second substrate is an amino acid sequence containing -leucine-glycine-glutamine-glycine-.
74. The cell of claim 51, wherein the second substrate is an amino acid sequence containing -leucine-glutamine-glycine-glycine-.
75. The cell of claim 51, wherein the second substrate is an amino acid sequence containing -leucine-leucine-glutamine-glycine-.

76. The cell of claim 51, wherein the second substrate is a modified staphylococcal nuclease ("SNase") or a modified thioredoxin containing an amino acid sequence containing a glutamine.
77. The cell of claim 51, wherein the enzyme is a transglutaminase.
78. The cell of claim 51, wherein the enzyme is a microbial transglutaminase, a tissue transglutaminase, or Factor XIIIa.
79. A kit for detecting bond formation by an enzyme between a first substrate and a second substrate in a cell, comprising
- (a) a host cell containing a reporter gene that is expressed only when bound to a DNA-binding domain and when in the proximity of a transcription activation domain;
 - (b) a first vector containing a promoter that functions in the host cell and a DNA encoding a DNA-binding domain;
 - (c) a second vector containing a promoter that functions in the host cell and a DNA encoding a transcription activation domain;
 - (d) a third vector containing a promoter that functions in the host cell;
 - (e) a dimeric small molecule which comprises a moiety known to bind a receptor domain and a moiety containing the first substrate of the enzyme;
 - (f) a means for inserting into the first vector or the second vector a DNA encoding a receptor domain in such a manner that the receptor domain and the DNA-binding domain are expressed as a fusion protein;
 - (g) a means for inserting into the first vector or the second vector a DNA encoding a protein containing the second substrate of the enzyme in such a manner that the protein and

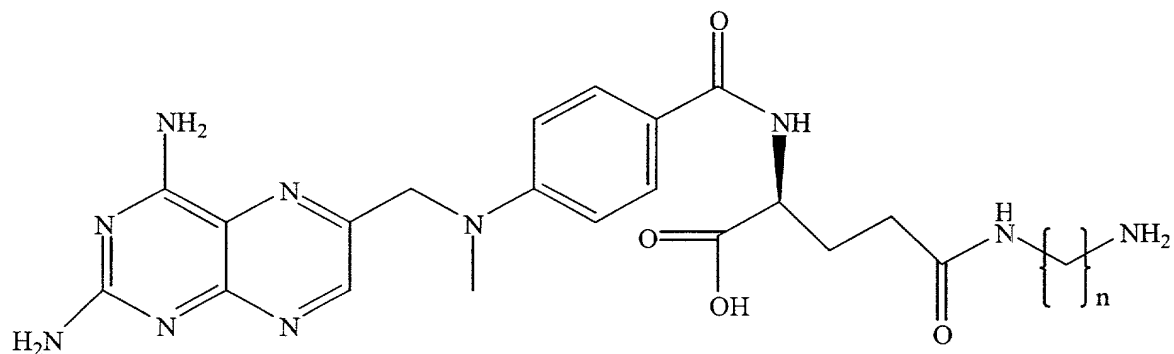
the transcription activation domain are expressed as a fusion protein;

(h) a means for inserting into the third vector a DNA encoding the enzyme; and

(h) a means for transfecting the host cell with the first vector, the second vector, and the third vector,

wherein bond formation by the enzyme between the first substrate and the second substrate results in a measurably greater expression of the reporter gene than in the absence of bond formation by the enzyme.

80. A small molecule compound having the structure:



wherein n is an integer from 1 to 20.

81. The compound of claim 80, wherein n is an integer from 2 to 12.
82. The compound of claim 80, wherein n is an integer from 3 to 9.
83. The compound of claim 80, wherein n is 5.